

Gall Anonimizacja

**system automatycznej anonimizacji
dokumentów**

wersja 2.4.2

© **SoftProdukt** 2019

Warszawa, grudzień 2019 r.

Wstęp	3
1. Typy anonimizowanych dokumentów	3
2. Instalacja i uruchomienie.	3
3. Zakres i sposób automatycznej anonimizacji.....	4
4. Konfiguracja automatycznej anonimizacji.....	5
5. Praca z systemem Gall.	7
6. Weryfikacja dokumentów po automatycznej anonimizacji.	7
7. Dodatkowe informacje.	8
8. Wymagania Systemowe	8
9. Pliki programu Gall.....	8
10. Dane kontaktowe.....	8
Załącznik - Weryfikacja i korekta dokumentu w edytorze MS Word v. 2007	9

Wstęp

Gall Anonimizacja (zwany dalej **Gall**) jest systemem automatycznej anonimizacji dokumentów tekstowych. Wprowadzono w nim specjalizowane mechanizmy tekstowe, co pozwoliło znacznie przyspieszyć proces anonimizacji dokumentów i bezpieczeństwo udostępnianych danych.

Gall daje możliwość kontroli i prezentacji wszystkich zamienianych fragmentów tekstu („co” na „co”), odwracania wybranych anonimizacji oraz dodawania nowych.

Gall pozwala użytkownikom dostosować zakres anonimizacji do własnych potrzeb.

Gall zapewnia pełne bezpieczeństwo przetwarzanych danych, bo działa na lokalnym komputerze i nie wymaga żadnych połączeń sieciowych.

Należy podkreślić, że system anonimizacji **Gall** podlega stałemu rozwojowi i zmianom wynikającym z prowadzonych prac badawczych, rozwoju technologii oraz uwag zgłaszanych przez użytkowników.

Uwaga: W pełni funkcjonalna wersja testowa posiada wszystkie mechanizmy anonimizacji wersji komercyjnej, ale jej działanie jest ograniczone do pierwszych 4000 znaków.

1. Typy anonimizowanych dokumentów

Obecna wersja Gall 2.4.2 obsługuje następujące formaty dokumentów:

- a) MS Word – format “docx” i “doc”
- b) MS Excel – format “xlsx” i “xls”
- c) Adobe Acrobat Reader – format „pdf”

Uwaga: Animizowane są tylko dokumenty tekstowe. Napisy w obiektach graficznych (zdjęcia, obrazki, itp.) nie są obsługiwane bezpośrednio przez system Gall. Dokumenty graficzne należy najpierw przetworzyć programem OCR do postaci tekstowej, dopiero potem można wykorzystać Galla do automatycznej anonimizacji.

2. Instalacja i uruchomienie.

- 2.1. Rozpakować plik instalacyjny **gall.zip** i wgrać do wybranego katalogu komputera z systemem MS Windows.
- 2.2. Uruchomić program **gall.exe**.

3. Zakres i sposób automatycznej anonimizacji.

W wersji **Gall 2.4.2** dostępna jest automatyczna anonimizacja następujących danych:

- a) Imiona i nazwiska są zastępowane inicjałami, np. Jan Nowak jest zastępowany przez J. N. (system rozpoznaje imiona, słowa pisane wielką literą w ich otoczeniu traktuje jako nazwiska).
- b) Nazwy ulic wraz z numerami domów i lokali są zastępowane pierwszymi literami z kropkami. Rozpoznawanie słów pisanych wielką literą przy słowach „ul.”, „Al.” itp. Uwzględniane są nazwy jedno-, dwu- i trójslówowe.
- c) Kody pocztowe są zastępowane przez [...].
- d) Nazwy większych miejscowości w Polsce (ponad 900) są zastępowane przez [...] lub przez pierwszą literę (zgodnie z ustawieniem w konfiguracji).
- e) Strony www są zastępowane przez [...].
- f) Adresy e-mail są zastępowane przez [...].
- g) Adresy IP są zastępowane przez [...].
- h) Numery telefoniczne są zastępowane przez [...].
- i) Numery NIP, REGON, KRS, PESEL są zastępowane przez [...].
- j) Numery takie jak ciągi znaków występujące obok takich słów jak: „nr”, „nr.”, „numer” oraz „znak”, „znk”, „faktura”, „działka”, „pozycja”, „ewid.” są zastępowane przez [...].
- k) Inne numery, które mają przynajmniej 5 znaków długości, zawierają co najmniej jedną cyfrę, co najmniej jedną literę i co najmniej jeden znak specjalny, taki jak „/” (ukośnik), „.” (kropka), „-” (myślnik) są zastępowane przez [...].
- l) Kwoty (liczby występujące obok słów „kwota”, „zł”, „PLN”, „EUR”, „euro”, „USD”, „\$”) są zastępowane przez [...].
- m) Dаты – do wyboru: anonimizowany jest tylko numer dnia lub cała data – są zastępowane przez [...].

W konfiguracji można ustawić które nawiasy mają być wykorzystywane przy zamianach z trzema kropkami. Do wyboru są dwie wersje : [...] lub (...) .

Ponadto istnieje możliwość automatycznej anonimizacji dowolnych konkretnych słów indywidualnie zdefiniowanych przez użytkownika, np. nazwy loginów czy serwerów komputerowych - patrz pkt 4 (Konfiguracja automatycznej anonimizacji).

Uwaga:

Ze względu na obiektywne uwarunkowania nie jest możliwa pełna automatyczna anonimizacja tekstu dokumentów w ramach wybranych reguł. Nawet dysponując wzorcami tekstów do anonimizacji (np. nazwisko), nie zawsze można 100% dokładnie rozpoznać wszystkie wystąpienia danego wzorca w tekście, z uwagi m.in. na fleksję języka polskiego czy literówki w oryginalnym dokumencie. Dlatego zawsze wstępna automatyczna anonimizacja musi podlegać weryfikacji, ewentualnej korekcie i zatwierdzeniu przez operatora.

4. Konfiguracja automatycznej anonimizacji.

Automatyczną anonimizację użytkownik może dostosowywać do własnych wymagań.

4.1. W menu Ustawienia

Istnieje możliwość konfigurowania automatycznej anonimizacji przez wybór z dostępnych algorytmów anonimizacji. W tym celu wystarczy odznaczyć lub zaznaczyć wybrane pozycje na poniższej liście (domyślnie wszystkie pozycje są zaznaczone).



Po zaakceptowaniu przyciskiem **Zapisz** nowa konfiguracja zostanie zapisana w pliku **gall.ini** i będzie nową obowiązującą konfiguracją domyślną

4.2. Słowniki użytkownika

Są słowa (zwroty), które chcemy automatycznie anonimizować, a które nie mogą być rozpoznane ww. algorytmami. Dotyczy to np. nazwisk, które występują samodzielnie (bez imion), loginów czy nazw serwerów (pozycje te mogą być praktycznie dowolnymi wyrazami). Takie słowa (zwroty) mogą być rozpoznawane na podstawie lokalnych słowników użytkownika.

Słowniki muszą być plikami tekstowymi i znajdować się w podkatalogu domowym systemu **Gall**. Słowa (zwroty) w słownikach

- w podkatalogu **../słowniki** będą zamieniane na **[...]** ,
- w podkatalogu **../słowniki_inicjały** będą zamieniane na **pierwsze litery (inicjały)**.

Słowniki mogą mieć dowolne nazwy, ale lepiej nazywać je zgodnie z ich zawartością, np. **nazwiska.txt**, **loginy.txt**, **serwery.txt** itd.

Słowniki mogą być tworzone na bieżąco, wypełniane i uzupełniane przez użytkowników stosownie do potrzeb, np. przy użyciu edytora MS Word lub notatnika MS Windows.

Słowa występujące w słownikach użytkownika należy wpisywać po jednym w wierszu. Będą rozpoznawane dokładnie, literalnie tak, jak zostały wpisane, tzn. z uwzględnieniem wielkości liter i bez uwzględnienia innych form w odmianie.

Jeśli na przykład zostanie wpisane tylko słowo „Kowalski”, to automatyczna anonimizacja nie rozpozna słów „kowalski”, Kowalskiemu”, „Kowalskiego”, „Kowlski” itd.

*Przykład zawartości pliku **loginy.txt***

admin
admin1
user1
redaktor23

4.3. Słowniki wyjątków

Są również słowa (zwroty) wyjątki, których nie chcemy anonimizować, chociaż są rozpoznawane przez algorytmy automatycznej anonimizacji. Może to dotyczyć np. imion i nazwisk autorów opracowań naukowych, osób publicznych lub innych nazw. Takie słowa (zwroty) mogą być wyłączone z automatycznej anonimizacji na podstawie słowników wyjątków. Słowniki wyjątków muszą być plikami tekstowymi i znajdować się w podkatalogu domowym **../wyjatki** systemu **Gall**. Budowa słowników i zasady rozpoznawania są identyczne jak dla słowników użytkownika opisanych powyżej.

*Przykład zawartości pliku z wyjątkami **nazwiska.txt** jeśli nie chcemy anonimizować autorów opracowań i słowników, które często występują w naszych dokumentach.*

Władysław Kopaliński
W. Kopaliński
Kazimierz Jaśkowski
K. Jaśkowski
Eliza Maniewska
E. Maniewska

Uwaga: Aby dany zwrot nie był anonimizowany musi być wpisany dokładnie w CAŁEJ postaci, która podlega anonimizacji. Na przykład: jeśli w tekście występuje zwrot „Władysław Kopaliński”, to jako osoba będzie on podlegał anonimizacji do postaci „W. K.”. Jeśli w słowniku wyjątków będzie wpisane samo nazwisko „Kopaliński”, to wyjątek „nie zadziała”, ponieważ zwrot „Kopaliński” nie odpowiada zwrotowi podlegającemu anonimizacji „Władysław Kopaliński”.

4.4. Słownik **anonimizuj_po**

Jeśli nie chcemy anonimizować początku tekstu, np. danych instytucji lub osób podejmującej decyzję, to można określić miejsce w dokumencie od którego **Gall** rozpocznie automatyczną anonimizację.

W tym celu w słowniku **anonimizuj_po** należy umieścić wybrane słowa (zwroty) kluczowe, np. „po rozpatrzeniu”, „po rozpoznaniu”. W takim przypadku **Gall** będzie anonimizował dokument od pierwszego wystąpienia dowolnego z tych słów. Jeśli w dokumencie nie wystąpi żadne ze słów (zwrotów) ze słownika, to anonimizowany będzie cały dokument.

Słownik **anonimizuj_po** znajduje się w katalogu domowym **Galla**. Budowa słownika i zasady rozpoznawania są identyczne jak dla słowników użytkownika opisanych powyżej.

*Przykład zawartości pliku **anonimizuj_po**.*

po rozpoznaniu
po rozpatrzeniu
na posiedzeniu
w sprawie
sprawy

5. Praca z systemem Gall.

Praca z systemem **Gall** jest intuicyjna i oparta na zasadach pracy w systemie MS Windows i MS Office.

Po wywołaniu programu **Gall** pojawia się następujący ekran



Anonimizacja plików.

- a) Wybrać (otworzyć) plik do anonimizacji.
Jako domyślny plik wynikowy zostaje pokazany plik z identyczną nazwą i z dopiskiem „**gall**” do zapisania w tym samym katalogu, co plik źródłowy.
Nazwę i lokalizację pliku można zmienić.
- b) W celu uruchomienia anonimizacji wybranego pliku nacisnąć „**Anonimizuj**”.
Uwaga: Jeśli plik o nazwie docelowej już istnieje, wtedy pojawi się komunikat „Czy nadpisać plik wynikowy (Tak / Nie) ?”. Wybór „Nie” przerwie anonimizację.
Pojawi się informacja „**Przetwarzanie ...**”, a następnie „**Zakończono przetwarzanie pliku <nazwa pliku źródłowego>**”.
- c) Można anonimizować kolejne dokumenty lub otworzyć plik wynikowy w celu weryfikacji (punkt 6 poniżej).

6. Weryfikacja dokumentów po automatycznej anonimizacji.

Zależnie od typu (docx, doc, pdf lub xlsx, xls) dokument należy otworzyć programem MS Word lub MS Excel.

Dokument zostaje otwarty w trybie rejestracji i pokazywania zmian (*).

Wszystkie wprowadzone przez system **Gall** zmiany są prezentowane na ekranie („co” na „co”)

Dalsza edycja jest już realizowana narzędziami MS Office.

Przeglądając dokument, można wycofać każdą znalezioną błędną zmianę lub „ręcznie” wprowadzić zmianę, która nie została wykonana automatycznie przez system **Gall**.

Zmiany zarejestrowane w dokumencie można zatwierdzać w trakcie przeglądania lub globalnie

po zweryfikowaniu dokumentu.

- (*) *W obecnej wersji, aby pokazać wprowadzone zmiany w pliku xls lub xlsx, po uruchomieniu programu Excel należy włączyć wyróżnianie zmian. W tym celu należy wybrać menu **Recenzja**, a następnie **Śledź zmiany i Wyróżnij zmiany**. Pojawi się ekran **Wyróżnianie zmian**. Należy na nim odznaczyć pozycję **Kiedy**. (Opis dla wersji Excel 2013. W innych wersjach Excela włączanie wyróżniania zmian może się różnić od powyższego opisu)*

7. Dodatkowe informacje.

7.1. Menu **Pomoc**

W menu **Pomoc** znajduje się wersja elektroniczna niniejszego podręcznika **Galla**.

7.2. Menu **O programie**

W menu **O programie** znajdują się informacje o wersji programu **Gall**, zasady korzystania z oprogramowania oraz serwisowe dane kontaktowe.

8. Wymagania Systemowe

8.1. MS Windows od wersji 7

8.2. MS Word od wersji 2007–anonimizacja plików docx i doc

Uwaga: anonimizacja dokumentów pdf od wersji MS Word 2013

8.3. MS Excel od wersji 2007–anonimizacja plików xlsx i xls

9. Pliki programu Gall

- **gall.exe** – aplikacja **Gall**
- **gall.wzr** – standardowe reguły anonimizacji
- **gall.ini** – domyślna konfiguracja
- **gall.lic** – licencja programu
- **../slovníki** – podkatalog zawierający dodatkowe słowniki użytkownika z zamianą na [...] (podkatalog musi istnieć, może być pusty)
- **../slovníki_inicjaly** – podkatalog zawierający dodatkowe słowniki użytkownika z zamianą na pierwsze litery/inicjały (podkatalog musi istnieć, może być pusty)
- **../wyjatki** – podkatalog zawierający dodatkowe słowniki użytkownika z wyjątkami (podkatalog musi istnieć, może być pusty)
- **anonimizuj_po.txt** – słownik ze słowami/zwrotami, które wyznaczają początek anonimizacji w tekście
- **pomoc.pdf** – instrukcja użytkownika (ten dokument)
- **licencja.pdf** – warunki licencji na system **Gall Anonimizacja**

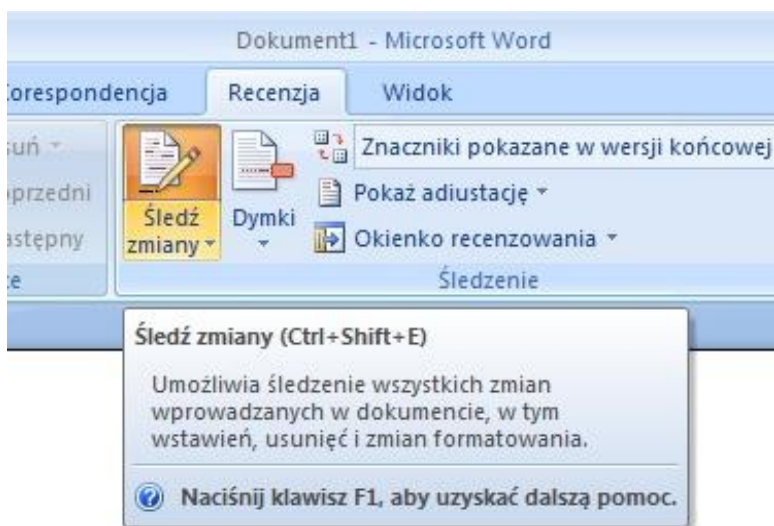
10. Dane kontaktowe

SoftProdukt Sp. z o.o.

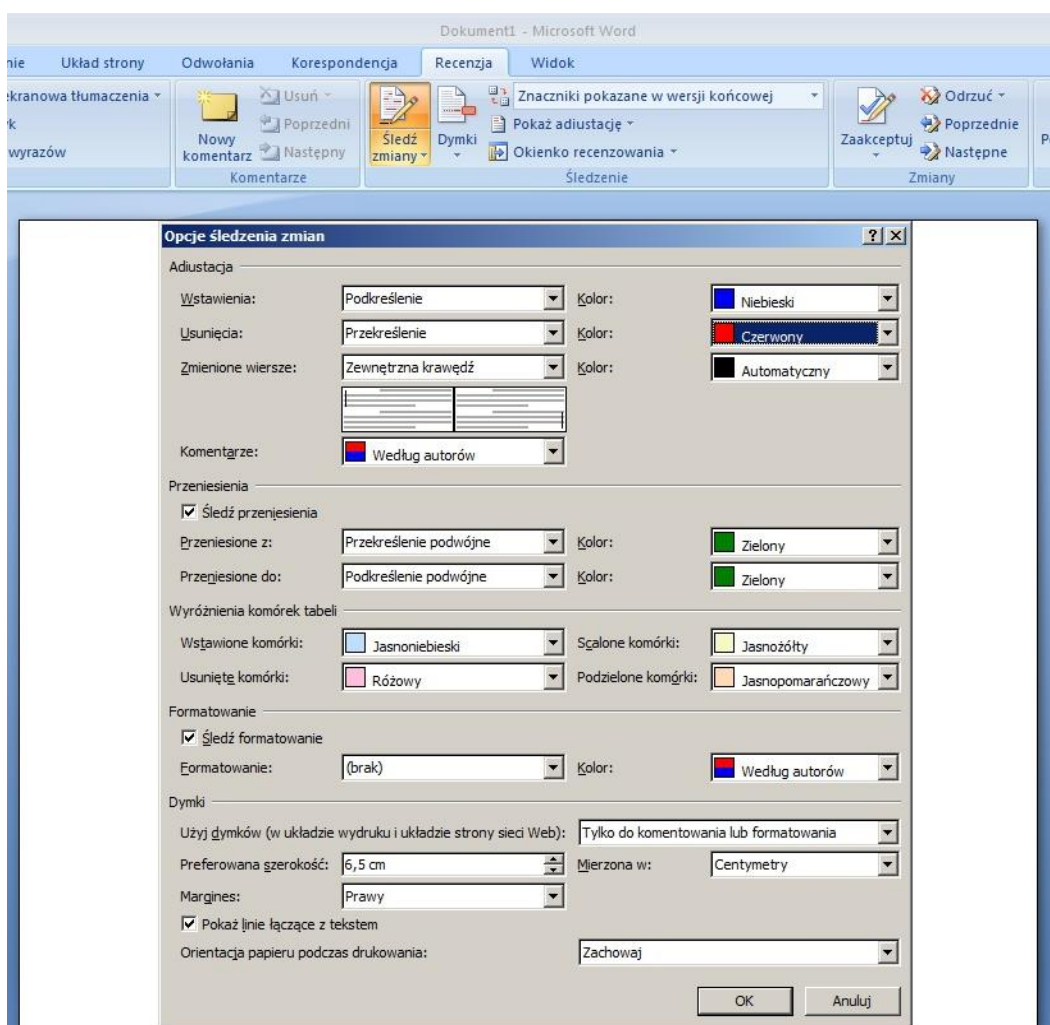
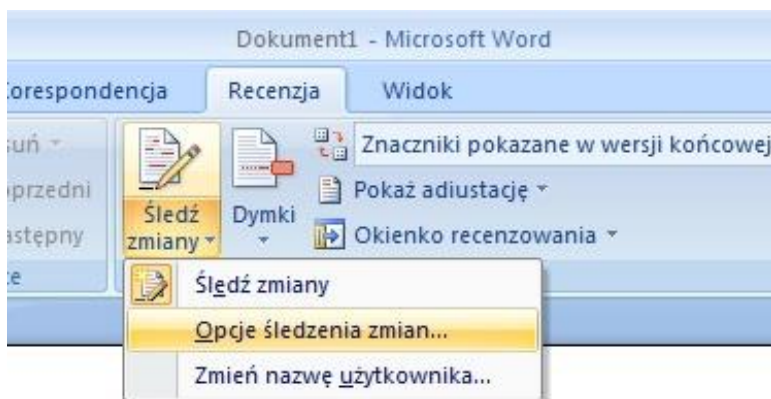
e-mail: gall@softprodukt.com.pl

Załącznik - Weryfikacja i korekta dokumentu w edytorze MS Word v. 2007

1. Włączenie śledzenia zmian (menu <Recenzja>)

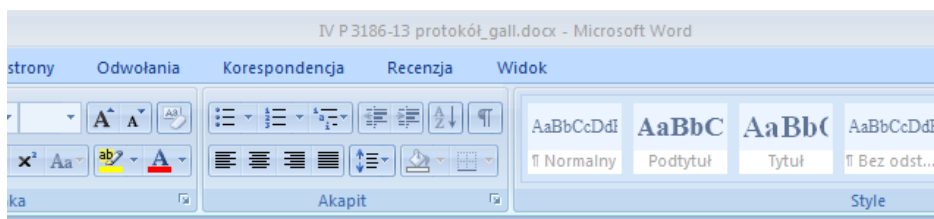


2. **Formatowanie zmian (menu <Recenzja>)** (np. usunięte – przekreślony czerwony, dodane – podkreślony niebieski)



3. Dodanie pojedynczej anonimizacji:

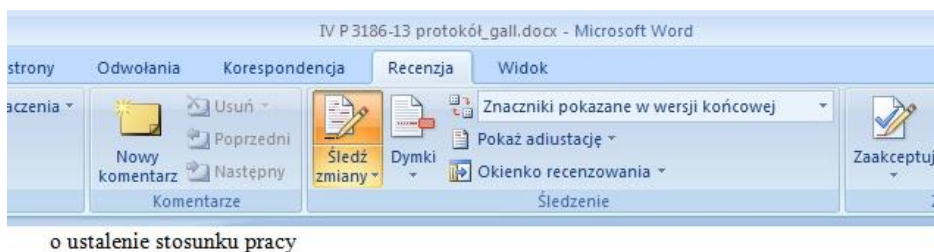
- zaznaczenie fragmentu,
- usunięcie (np. klawisz),
- wstawienie anonimizacji, np. inicjał z kropką lub [...] albo (...)



Posiedzenie rozpoczęło o godzinie 9:00 Posiedzenie zakończono o godzinie 9:40

Po wywołaniu sprawy stawili się:

S. P. – starszy inspektor pracy OIP w Zielonej Górze Oddział w G.5.



o ustalenie stosunku pracy

Posiedzenie rozpoczęło o godzinie 9:00 Posiedzenie zakończono o godzinie 9:40

Po wywołaniu sprawy stawili się:

S. P. – starszy inspektor pracy OIP w Zielonej Górze Z. Oddział w Gorzowie Wilkop G..

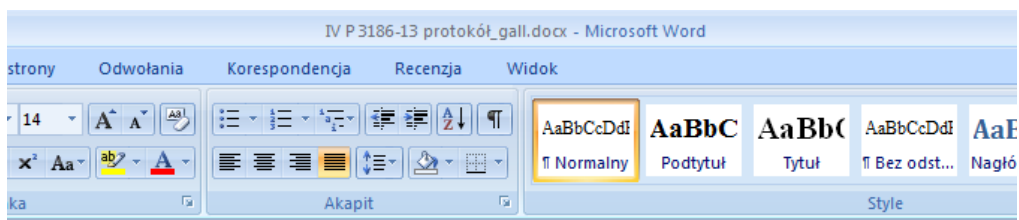
Piotr Rzyński P. R. - nie stawiał się, zawiadomiony prawidłowo

Za pomoc w przygotowaniu sprawdził w systemie Złoty T.

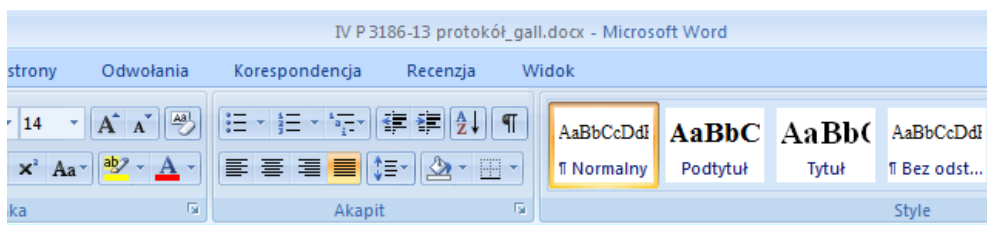
4. Dodanie automatyczne kolejnych anonimizacji (powtórzenia) np. dla firmy

„Eurokej”.

funkcja <Zamień>: Co „Eurokej”, na co: „E.”, <Zamień wszystko>. Nastąpi zamiana w całym dokumencie i powrót do miejsca gdzie był kursor przed wywołaniem funkcji zamiany.

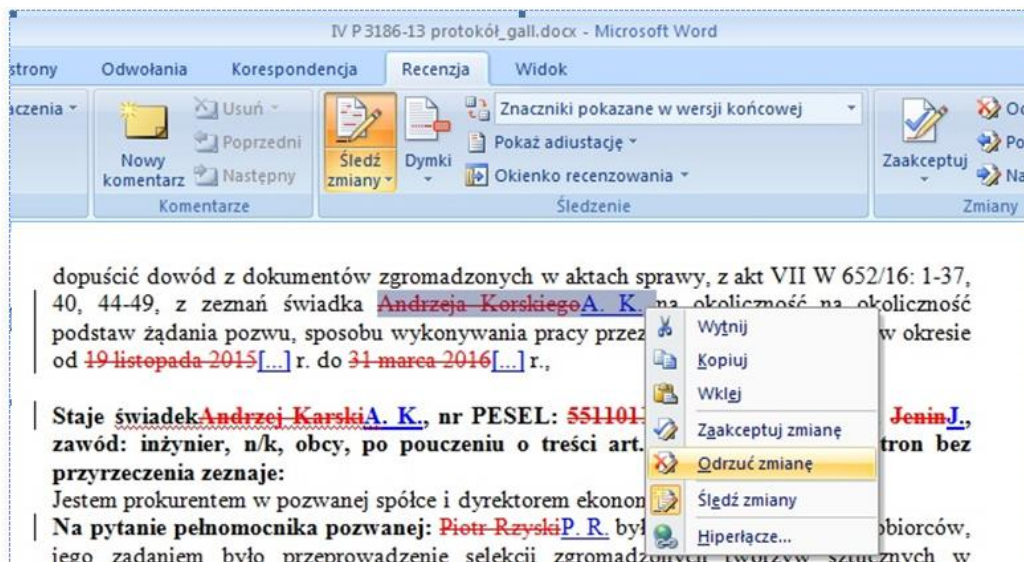


My wynajmowaliśmy pomieszczenia od firmy Link Plast i pracownicy wykonywali czynności na terenie Link Plast w Międzyrzeczu M.. Maszyny, sprzęt i urządzenia, które dostarczył Eurokej. Pracownicy Secret Service pracowali przy taśmie i na stole selekcjonując dostarczony materiał. Zlecenie obejmowało wyłącznie preselekcjonowanie dostarczonego materiału. Obsługę techniczną urządzeń miał zagwarantować Eurokej. My płaciliśmy za efektywny czas pracy. Nie było stosowanych norm dotyczących ilości wyselekcjonowanego materiału. Zleceniobiorcy skarżyli się, że pracownicy funkcyjni Eurokej i Link Plastu narzucali im dodatkowe obowiązki różnego

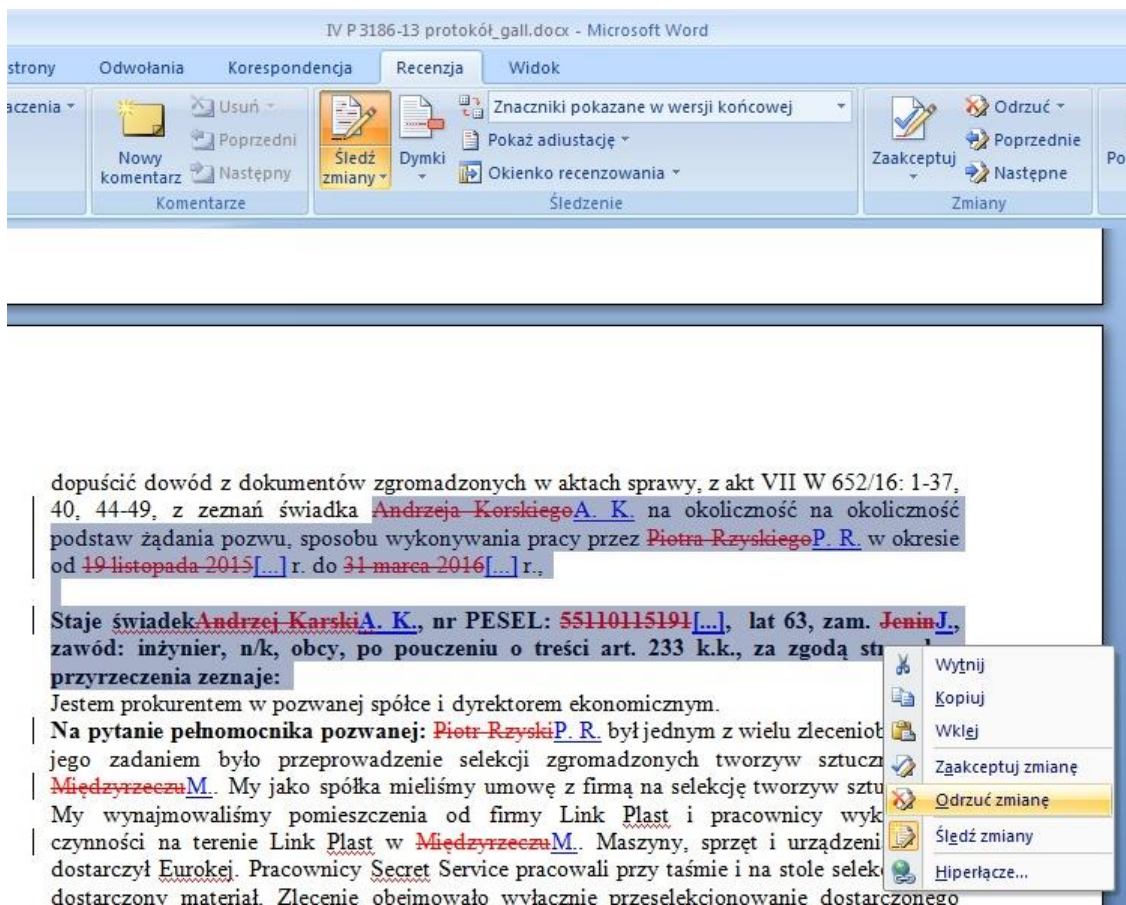



My wynajmowaliśmy pomieszczenia od firmy Link Plast i pracownicy wykonywali czynności na terenie Link Plast w Międzyrzeczu M.. Maszyny, sprzęt i urządzenia, które dostarczył Eurokej... Pracownicy Secret Service pracowali przy taśmie i na stole selekcjonując dostarczony materiał. Zlecenie obejmowało wyłącznie preselekcjonowanie dostarczonego materiału. Obsługę techniczną urządzeń miał zagwarantować Eurokej... My płaciliśmy za efektywny czas pracy. Nie było stosowanych norm dotyczących ilości wyselekcjonowanego materiału. Zleceniobiorcy skarżyli się, że pracownicy funkcyjni Eurokej... i Link Plastu narzucali im dodatkowe obowiązki różnego

5. **Wycofanie pojedynczej anonimizacji:** łączne zaznaczenie usuniętego i wstawionego fragmentu, prawy klawisz myszy i wybór <Odrzuć zmianę>

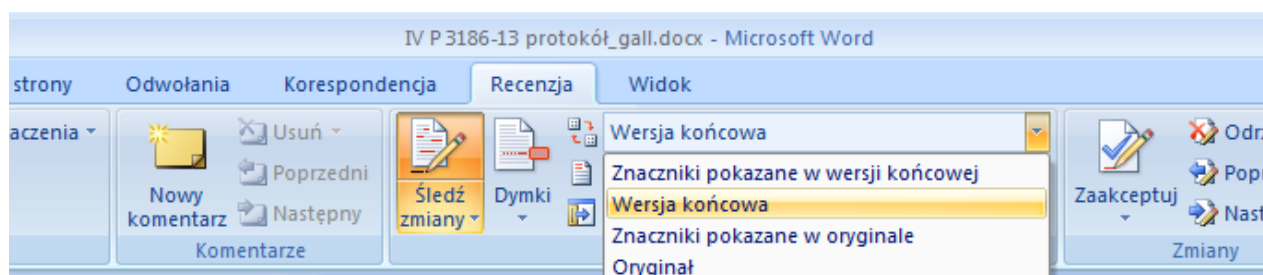


6. **Jednoczesne wycofanie wielu (grupy) anonimizacji:** zaznaczenie większego tekstu, zaczynając od pierwszej anonimizacji, zawierającego anonimizację, prawy klawisz myszy i wybór <Odrzuć zmianę>



7. **Wycofanie błędnie wprowadzonych zmian** - akceptacji anonimizacji, odrzucenia anonimizacji lub innych operacji pisania / usuwania: jedno- lub wielokrotne kliknięcie guzika „cofnij”  w górnym pasku z lewej strony.

8. Widok dokumentu po zmianach (menu <Recenzja>)



Po wywołaniu sprawy stawili się:

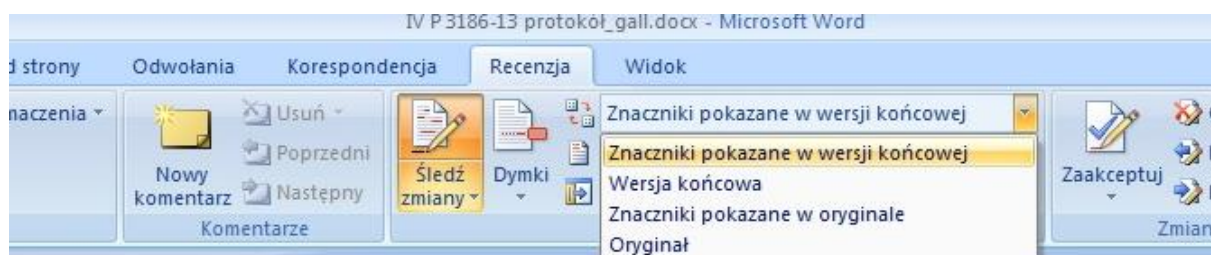
S. P. – starszy inspektor pracy OIP w Z. Oddział w G..

P. R. - nie stawiał się, zawiadomiony prawidłowo

Za pozwaną spółkę pełnomocnik r.pr. J. Ż.

Stawił się świadek A. K.

9. Przywrócenie widoku ze zmianami (menu <Recenzja>)



Po wywołaniu sprawy stawili się:

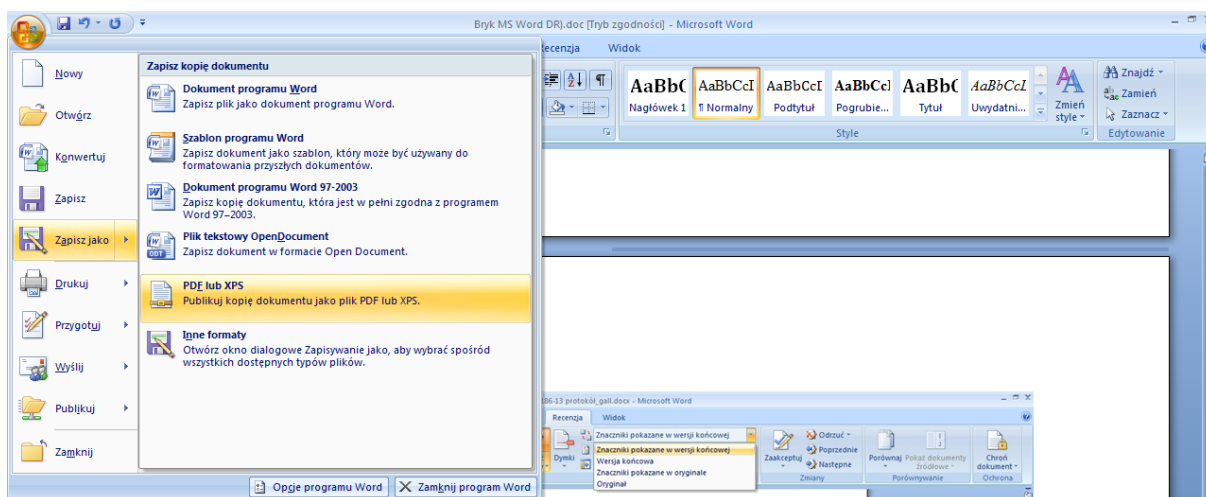
~~Szczepan Przybylak~~ S. P. – starszy inspektor pracy OIP w ~~Zielonej Górze~~ Z. Oddział w ~~Gorzowie Wlkp~~ G..

~~Piotr Rzyński~~ P. R. - nie stawiał się, zawiadomiony prawidłowo

Za pozwaną spółkę pełnomocnik r.pr. ~~Justyna Żółta~~ J. Ż.

Stawił się świadek ~~Andrzej Karński~~ A. K.

10. Wydruk (zapisanie) dokumentu do formatu PDF (widok po zmianach – weryfikacja ustawienia drukowania w ustawieniach menu).



11. Zachowanie roboczej wersji dokumentu wordowego do ewentualnych dalszych korekt. Nie wywołujemy funkcji zatwierdzenia wszystkich zmian.

